IN THE UNITED STATES PATENT AND TRADEMARK OFFICE
BEFORE THE BOARD OF PATENT APPEALS AND INTERFERENCES

First Named Inventor : Dong Yu et al.	
Appln. No.: 10/761,451	Confirmation No.: 3046
Filed : January 20, 2004	Group Art Unit: 2626
For : AUTOMATIC SPEECH RECOGNITION LEARNING USING USER CORRECTIONS	Examiner: Shah
Docket No.: M61.12-0582	

BRIEF FOR APPELLANTS

Mail Stop Appeal Brief-Patents Commissioner for Patents P.O. Box 1450 Alexandria, VA 22313-1450	**Electronically Filed on April 11, 2008**

Sir:

This is an appeal from a Final Office Action mailed November 7, 2007 in which claims 1, 3, 4, 6-9 and 11-25 were rejected. Appellants respectfully submit that claims 1, 3, 4, 6, 7, and 14-25 are allowable, and request that the Board reverse the rejection of claims 1, 3, 4, 6, 7, and 14-25 and find that claims 1, 3, 4, 6, 7, and 14-25 are in condition for allowance.

REAL PARTY IN INTEREST

Microsoft Corporation, a corporation organized under the laws of the state of Washington, and having a place of business at One Microsoft Way, Redmond, WA, 98052, has acquired the entire right, title and interest in and to the invention, the application, and any and all patents to be obtained therefor, as set forth in the Assignment filed with the patent application and recorded on Reel 014915, Frame 0119.

NO RELATED APPEALS OR INTERFERENCES

There are no known related appeals or interferences which will directly affect or be directly affected by or have a bearing on the Board's decision in this appeal.

STATUS OF THE CLAIMS

Claims 1-22 were originally presented. Claims 1, 4, 7-9, 11 and 14 were amended; claims 2, 5 and 10 were cancelled; and new claims 23-25 were added on September 24, 2007. Claims 7, 14, 15, 21 and 22 were amended and claims 8, 9 and 11-13 were cancelled by way of an Amendment After Final filed on January 7, 2008. Appellants have been informed that the Amendment After final will be entered. Thus, the pending and rejected claims 1, 3, 4, 6, 7 and 14-25 are the subject of the present appeal.

STATUS OF AMENDMENTS

An Amendment After Final was filed on January 7, 2008. Appellants have been informed that the Amendment After Final will be entered.

SUMMARY OF CLAIMED SUBJECT MATTER

1. Introduction

The present invention relates to computer speech recognition.

2. Brief Background

The rapid and accurate recognition of human speech by a computer system has been a long-sought goal by developers of computer systems. The benefits that would result from such a computer speech recognition (CSR) system are substantial. For example, rather than typing a document into a computer system, a person could simply speak the words of the document, and the CSR system would recognize the words and store the letters of each word as if the words had been typed. Since people generally can speak faster than type, efficiency would be improved. Also, people would no longer need to learn how to type. Computers could also be used in many applications where their use is currently impracticable because a person's hands are occupied with tasks other than typing.

Typical CSR systems recognize words by comparing a spoken utterance to a model of each word in a vocabulary. The word whose model best matches the utterance is recognized as the spoken word. A CSR system may model each word as a sequence of phonemes that compose the word. To recognize an utterance, the CSR system identifies a word sequence, the phonemes of which best match the utterance. These phonemes may, however, not exactly correspond to the phonemes that compose a word. Thus, CSR systems typically use a probability analysis to determine which word most closely corresponds to the identified phonemes.

When recognizing an utterance, a CSR system converts the analog signal representing the utterance to a more useable form for further processing. The CSR system first converts the analog signal into a digital form. The CSR system then applies a signal processing technique, such as fast fourier transforms (FFT), linear predictive coding (LPC), or filter banks, to the digital form to extract an appropriate parametric representation of the utterance. A commonly used representation is a "feature vector" with FFT or LPC coefficients that represent

the frequency and/or energy bands of the utterance at various intervals (referred to as "frames"). The intervals can be short or long based on the computational capacity of the computer system and the desired accuracy of the recognition process. Typical intervals may be in the range of 10 milliseconds. That is, the CSR system would generate a feature vector for every 10 milliseconds of the utterance. Each frame is typically 25 ms long. Therefore, a 25 ms long frame is generated every 10 ms. There is an overlap between successive frames.

To facilitate the processing of the feature vectors, each feature vector is quantized into one of a limited number (e.g., 256) of "quantization vectors." That is, the CSR system defines a number of quantization vectors that are selected to represent typical or average ranges of feature vectors. The CSR system then compares each feature vector to each of the quantization vectors and selects the quantization vector that most closely resembles the feature vector to represent the feature vector. Each quantization vector is uniquely identified by a number (e.g., between 1 and 256), which is referred to as a "codeword." When a feature vector is represented as a quantization vector, there is a loss of information because many different feature vectors map to the same quantization vector. To ensure that this information loss will not seriously impact recognition, CSR systems may define thousands or millions of quantization vectors. The amount of storage needed to store the definition of such a large number of quantization vectors can be considerable. Thus, to reduce the amount of storage needed, CSR systems segment feature vectors and quantize each segment into one of a small number (e.g., 256) quantization vectors. Thus, each feature vector is represented by a quantization vector (identified by a codeword) for each segment. For simplicity of explanation, a CSR system that does not segment a feature vector and thus has only one codeword per feature vector (or frame) is described.

As discussed above, a spoken utterance often does not exactly correspond to a model of a word. The difficulty in finding an exact correspondence is due to the great variation in speech that is not completely and accurately captured by the word models. These variations result from, for example, the accent of the speaker, the speed and pitch at which a person speaks, the current health (e.g., with a cold) of the speaker, the age and sex of the speaker, etc. CSR systems that use probabilistic techniques have been more successful in accurately recognizing speech than

techniques that seek an exact correspondence.

One such probabilistic technique that is commonly used for speech recognition is hidden Markov modeling. A CSR system may use a hidden Markov model ("HMM") for each word in the vocabulary. The HMM for a word includes probabilistic information from which can be derived the probability that any sequence of codewords corresponds to that word. Thus, to recognize an utterance, a CSR system converts the utterance to a sequence of codewords and then uses the HMM for each word to determine the probability that the word corresponds to the utterance. The CSR system recognizes the utterance as the word with the highest probability.

An HMM is represented by a state diagram. State diagrams are traditionally used to determine a state that a system will be in after receiving a sequence of inputs. A state diagram comprises states and transitions between source and destination states. Each transition has associated with it an input which indicates that when the system receives that input and it is in the source state, the system will transition to the destination state. Such a state diagram could, for example, be used by a system that recognizes each sequence of codewords that compose the words in a vocabulary. As the system processes each codeword, the system determines the next state based on the current state and the codeword being processed. In this example, the state diagram would have a certain final state that corresponds to each word. However, if multiple pronunciations of a word are represented, then each word may have multiple final states. If after processing the codewords, the system is in a final state that corresponds to a word, then that sequence of codewords would be recognized as the word of the final state.

An HMM, however, has a probability associated with each transition from one state to another for each codeword. For example, if an HMM is in state 2, then the probability may be 0.1 that a certain codeword would cause a transition from the current state to a next state, and the probability may be 0.2 that the same codeword would cause a transition from the current state to a different next state. Similarly, the probability may be 0.01 that a different codeword would cause a transition from the current state to a next state. Since an HMM has probabilities associated with its state diagram, the determination of the final state for a given sequence of codewords can only be expressed in terms of probabilities. Thus, to determine the probability of

each possible final state for a sequence of codewords, each possible sequence of states for the state diagram of the HMM needs to be identified and the associated probabilities need to be calculated. Each such sequence of states is referred to as a state path.

To determine the probability that a sequence of codewords represents a phoneme, the CSR system may generate a probability lattice. The probability lattice for the HMM of a phoneme represents a calculation of the probabilities for each possible state path for the sequence of codewords. The probability lattice contains a node for each possible state that the HMM can be in for each codeword in the sequence. Each node contains the accumulated probability that the codewords processed so far will result in the HMM being in the state associated with that node. The sum of the probabilities in the nodes for a particular codeword indicates the likelihood that the codewords processed so far represent a prefix portion of the phoneme.

The accuracy of a CSR system depends, in part, on the accuracy of the output and transition probabilities of the HMM for each phoneme. Typical CSR systems "train" the CSR system so that the output and transition probabilities accurately reflect speech of the average speaker. During training, the CSR system gathers codeword sequences from various speakers for a large variety of words. The words are selected so that each phoneme is spoken a large number of times. From these codeword sequences, the CSR system calculates output and transition probabilities for each HMM. Various iterative approaches for calculating these probabilities are well-known.

A problem with such training techniques, however, is that such average HMMs may not accurately model the speech of people whose speech pattern is different than the average. In general, every person will have certain speech patterns that differ from the average. Consequently, CSR systems allow a speaker to train the HMMs to adapt to the speaker's speech patterns. In such training, CSR systems refine the HMM parameters, such as the output and transition probabilities and the quantization vectors represented by the codewords, by using training utterances spoken by the actual user of the system. The adapted parameters are derived by using both the user-supplied data as well as the information and parameters generated from the large amount of speaker-independent data. Thus, the probabilities reflect speaker-dependent

characteristics

A CSR system is typically trained by presenting a large variety of pre-selected words to a speaker. These words are selected to ensure that a representative sample of speech corresponding to each phoneme can be collected. With this representative sample, the CSR system can ensure that any HMM that does not accurately reflect the speaker's pronunciation of that phoneme can be adequately adapted. Since the CSR system functions in terms of probabilities, the more training that is provided, the more accurate subsequent speech recognition will be. However, as more and more training is done, the degree to which recognition accuracy will increase for a given amount of additional training begins to decline. Further, requiring user's to provide substantial investments in training time may diminish the user's experience.

Accordingly, there is a balance between the degree to which the user is called upon to train the system, and the degree to which the user can effectively use the system. Given the complexities of human language, it is very conceivable that even after extensive training, the system will occasionally generate errors. Another reason that causes a spoken utterance to not be matched with a corresponding model of a word, is when the word is new. A possible solution includes increasing the vocabulary size, which may lower recognition accuracy. Another solution is through user training in which the user adds new words. Current systems allow users to manually add new words with his or her pronunciation to a suitable lexicon, whether it be a system lexicon, a vendor or application lexicon, or a user-specific lexicon by using a user interface that allows a user to add or delete a word like an ADD/DELETE Words Dialog box. However, this can become troublesome in cases where users may need to add a significant number of words. It is also known to adapt the language model (LM) using documents and e-mails authored by the user. This approach is limited in that pronunciations are not added into the lexicon and the quality of the language model adaptation depends largely on the filtering of the source documents.

Thus, a need exists for a system that can easily learn new words and pronunciations thereof from users without requiring significant user intervention. Achieving this object would allow enhanced automatic speech recognition system learning without diminishing

the user experience by requiring undue training effort.

3. The Present Invention

Claims 1, 7 and 24 are the only independent claims on appeal.

Claim 1 provides a computer-implemented speech recognition system. A microphone (illustrated in Fig. 1 at reference numeral 163 and described on page 15, lines 9-12) receives user speech. A speech recognition engine is coupled to the microphone and recognizes the user speech and provides a textual output on a user interface. The system is adapted to recognize a user changing the textual output (as described on page 19, lines 6-13) and automatically, selectively adapt the speech recognition engine to learn from the change. The recognition engine is adapted to determine if a user's pronunciation caused an error (as described on page 19, line 13 – page 20, line 12), and selectively modify a probability associated with an existing pronunciation (as described on page 21, lines 21-24).

Claim 7 provides a method of learning with an automatic speech recognition system. The method includes detecting a change to dictated text (as described on page 19, lines 6-13) and inferring whether the change is a correction, or editing. Inferring whether the change is a correction, or editing includes comparing a speech recognition engine score of the dictated text and of the changed text (as described on page 20, lines 3-8). If the change is inferred to be a correction, the method selectively learns from the nature of the correction without additional user interaction (as described on page 21, lines 10-24). The selective learning from the nature of the correction includes determining if the corrected word exists in the user's lexicon (as described on page 20, lines 14-15), and if the corrected word does exist in the user lexicon, selectively learning the pronunciation (as described on page 20, line 24 – page 21, line 5).

Claim 24 provides a method of learning with an automatic speech recognition system. The method includes detecting a change to dictated text (as described on page 19, lines 6-13) and inferring whether the change is a correction based at least partially upon the number of words changed (as described on page 19, lines 25-28). If the change is inferred to be a correction,

the method selectively learns from the nature of the correction (as described on page 21, lines 10-24).

GROUNDS OF REJECTION TO BE REVIEWED ON APPEAL

Whether claims 1, 4 and 7 are anticipated by Nassiff et al. (U.S. Patent No. 6,418,416 – hereinafter Nassiff);

Whether claims 7, 21 and 22 are obvious over Nassiff in view of Hon et al. (U.S. Patent No. 5,852,801 – hereinafter Hon '801);

Whether claims 14-20 are obvious over Nassiff in view of Hon '801 and further in view of Hon et al. (U.S. Patent No. 5,963,903 – hereinafter Hon '903); and

Whether claims 24 and 25 are obvious over Nassiff in view of Gould (EP 0773532 A2).

Appellants respectfully submit that claims 1, 3, 4, 6, 7 and 14-25 are patentable over these references, and request that the Board find likewise and accordingly reverse the rejection of claims 1, 3, 4, 6, 7 and 14-25 and find these claims allowable.

ARGUMENT

1. Introduction: Claims 1, 3, 4, 6, 7 and 14-25 Should Be Allowed

With this appeal, the appellants respectfully request that the Board reverse the rejection of claims 1, 3, 4, 6, 7 and 14-25. As Appellants will explain below, the primary reference used to reject all of the claims (Nassiff) has been misconstrued and does not actually teach that which the Examiner asserts.

2. Anticipation

As set forth in Section Five of the Final Office Action, 35 U.S.C. §102(b) provides that a person shall be entitled to a patent unless, "(b) the invention was patented or described in a printed publication in this or a foreign country or in public use or on sale in this country, more than one year prior to the date of application for patent in the United States." Further, Appellants respectfully note that the Federal Circuit has provided guidance with respect to anticipation. The Federal Circuit has held anticipation to be present, "If every limitation in a claim is found in a single prior art reference." See Nystrom v. Trex Co., 71 U.S.P.Q.2.d. 1241 (Fed. Cir. 2004). Appellants respectfully submit that each and every element of independent claims 1 and 7 is not found in the Nassiff reference.

2.1 Claims 1 and 4 are Not Anticipated by Nassiff

Section Six of the Final Office Action indicated that independent claim 1, among others, was rejected under 35 U.S.C. §102(b) as being anticipated by Nassiff. With respect to this rejection, the Office Action asserted, on page 7, that Nassiff teaches,

> "Wherein the recognition engine is adapted to determine if the user's pronunciation caused the error, and selectively modify a probability associated with an existing pronunciation (see col. 7, lines 55-66) (e.g. The use of a statistical quantity with the updating of a language model implies that a probability value is associated with a word when comparisons are made (see col. 6, lines 28-31))."

Respectfully, the Office Action confuses the distinction between a probability associated with a word, such as may be present in a language model, and a probability associated with an existing pronunciation, as recited in independent claim 1. In this regard, Nassiff provides, in column 6, lines 28-32,

> "As is known by those skilled in the art, it should be understood that the language model consists of statistical information about word patterns. Accordingly, correcting the language model is not an acoustic correction, but a statistical correction." (Emphasis Added)

The quoted passage, as well as the rest of the Nassiff reference, teaches that the language model consists of statistical information about word patterns, and that the model accuracy can be improved by updating statistical information associated with word patterns. The claim 1

limitation recites modifying a probability associated with an existing pronunciation. Nassiff does not teach or suggest modification of a probability with respect to an existing pronunciation. Accordingly, Appellants respectfully submit that independent claim 1 is neither taught nor suggested by Nassiff.

In the Advisory Action mailed February 4, 2008, Page Three responded to Appellants' explanation of the distinction between modifying a probability associated with an existing pronunciation (as set forth in independent claim 1) and the updating of a language model (as Nassiff teaches). The Advisory Action cited column 6, lines 64-65 and column 7, lines 43-61 as allegedly showing "the updating of the language model and the relevance statistical scores (e.g. probability)." Appellants readily agree that Nassiff updates a "language model" and column 6, lines 64-65 and column 7, lines 43-61 confirm that. However, neither of those cited passages, nor the entire disclosure of Nassiff teaches or suggests the modification of a probability associated with an existing pronunciation when a user's pronunciation caused an error as set forth in independent claim 1. Appellants note that the Advisory Action continues providing,

> "Furthermore, the word patters as disclosed to Nassiff is a representation of word sequences (see col. 6, lines 60-66) that consists of probabilities associated with each other. A change in the sequence of word[s] directly affects the pronunciation, where the stated reference prevents future misrecognition by updating the language model (see col. 6, lines 33-34). The language model is updated in order to recognize words that may sound similar by modifying the probability (see col. 6, lines 45-50). The recognition of the correct dictation is updated and by this statistical correction the correct pronunciation dictated by the user is accepted so the error does not occur again between the words step and steep."

It appears that the Advisory Action asserts that the updating of the language model somehow changes the sequence of words, and that a changed sequence of words affects the pronunciation. Quite clearly, rearranging the words in a sentence will change the way sentence is pronounced. However, the clearly strained argument of the Advisory Action fails to address the fact that independent claim 1 recites "an existing pronunciation." Thus, for the argument of the Advisory Action to be relevant, the "sequence of word" recited to be changed must be an existing pronunciation. There is no indication in the reference, nor any logical explanation by the Advisory

Action, that this is the case. Accordingly, it is quite clear that Nassiff updates a language model while independent claim 1 modifies a probability associated with an existing pronunciation. These are quite different. Thus, Appellants respectfully submit that claims 1 and 4 are not anticipated by Nassiff.

2.2 Claim 7 Is Not Anticipated By Nassiff

Section Six of the Final Office Action also indicated that independent claim 7 was rejected under 35 U.S.C. §102(b) as being anticipated by Nassiff. Appellants have amended independent claim 7 by way of the Amendment After Final filed January 7, 2008 to recite the subject matter previously set forth in dependent claims 12 and 13. Accordingly, Appellants respectfully submit that the rejection of claim 7 recited in Section Six has been overcome. Appellants respectfully note that the subject matter previously set forth in dependent claims 12 and 13 was rejected in Section Eight of the Final Office Action under 35 U.S.C. §103(a) as being unpatentable over Nassiff in view of Hon '801. Appellants will address the obvious rejection later, but respectfully note that the rejection of those claims in Section Eight relies on the same construction of the Nassiff reference, which construction does not accurately reflect the distinction between updating a language model and updating a pronunciation. Claim 7 now recites selectively learning the pronunciation. This is in distinct contrast to Nassiff, which updates the language model. Accordingly, Appellants respectfully submit that amended independent claim 7 is allowable over Nassiff and Hon '801, taken alone or in combination.

3. The Law of Obviousness

To determine whether a claim is obvious, the scope and contents of the prior art at the time the invention was made must first be determined. *Graham v. John Deere*, 148 USPQ 459 (S.Ct. 1966).

Once the prior art is properly defined, the differences between the claimed invention as a whole and the prior art as a whole are evaluated. *Graham v. John Deere*; *Hodosh*

v. Block Drug Co., Inc., 229 USPQ 182 (Fed. Cir. 1986)(Rich, C.J.). This first requires construing the claims, according to the broadest reasonable meaning that the claim language would have to a person of ordinary skill in the art at the time the invention was made. *Phillips v. AWH Corp.*, 75 USPQ2d 1321 (Fed. Cir. 2005)(en banc)(Mayer, J. and Newman, J., dissenting). The test is not whether the individual differences themselves would have been obvious, but whether the claimed invention as a whole would have been obvious or not. *Stratoflex, Inc. v. Aeroquip Corp.*, 218 USPQ 871 (Fed. Cir. 1983).

3.1 Independent Claim 7 is Allowable Over Nassiff in view of Hon '801

With Appellants' Amendment After Final filed January 7, 2008, claim 7 now recites, "wherein selectively learning from the nature of the correction includes determining if the corrected word exists in the user's lexicon, and if the corrected word does exist in the user lexicon, selectively learning the pronunciation." Section Eight of the Final Office Action asserts that,

> "Hon et al. (801) does teach the use of a lexicon, which is updated for new words (see col. 9, lines 36-40), where words are added when determining if the words exist in the user lexicon (see col. 7, lines 66-67 and col. 8, lines 1-3) (e.g. The determination is made of whether the word is in the lexicon if it is unrecognized)."

Aside from being logically inconsistent, the above-quoted portion of the Final Office Action does indicate that Hon '801 is being applied where a word is a "new word" and thus would not exist in the user lexicon. In distinct contrast, independent claim 7 recites selectively learning the pronunciation if the corrected word "does" exist in the user lexicon. Accordingly, Appellants respectfully submit that the rejection of independent claim 7 under 35 U.S.C. §103 relying upon the Nassiff/Hon '801 combination fails to reach the subject matter of independent claim 7 not only because Nassiff does not update pronunciations (as set forth above) but because Hon '801 does not teach or suggest determining if the corrected word exists in the user's lexicon, and if the corrected word "does" exist in the user lexicon, selectively learning the

pronunciation. Accordingly, Appellants respectfully submit that independent claim 7 is allowable over Nassiff and Hon '801, taken alone or in combination.

3.2 Claims 14-20 are Allowable Over Nassiff in view of Hon '801 and Further in view of Hon '903

As an initial matter, Appellants respectfully submit that all of claims 14-20 depend, either directly or indirectly, from independent claim 7. Accordingly, the distinction set forth above with respect to the failing of the Nassiff reference to modify or learn a pronunciation applies equally to claims 14-20. Further, the failing of the Hon '801 reference to selectively learn the pronunciation when the corrected word "does" exist in the user lexicon also applies equally with respect to claims 14-20.

With respect to claim 14, specifically, the Final Office Action alleged that Hon '903 teach the aligning of waves based on a mis-recognized word and a correct word in column 6, lines 57-65 and column 7, lines 15-18. However, column 6, lines 57-65 merely provides,

> "The training system aligns a sequence of codewords with the phonemes of a word by first generating a probability lattice for the codewords and the known word. The training system then identifies the most-probable state path that leads to the most-probable state. The identification of such a state path preferably uses a Viterbi-based algorithm. The training system then uses the state path to identify which codewords would be recognized as part of (aligned with) which phonemes."

Further, Appellants respectfully note that "codeword" is defined by Hon '903 in column 1, lines 64-66. Specifically, "Each quantization vector is uniquely identified by a number (e.g., between 1 and 256), which is referred to as a "codeword." Accordingly, Appellants respectfully submit that column 6, lines 57-65 of Hon '903 does not teach or suggest the forced alignment of the wave based on a context word as recited in dependent claim 14. The other portion of Hon '903 cited by the Final Office Action is column 7, lines 15-18. However, that merely provides, "The determination of which phoneme models are less accurately modeled can be done by comparing the phoneme alignment of the utterance and misrecognized word against the phoneme alignment of the correct word." This cited portion of Hon '903 merely discusses

comparing alignments of the utterance and misrecognized word against the alignment of the correct word. It does not mention forcing an alignment of a wave based on a context word. Accordingly, Appellants respectfully submit that Hon '903 does not provide the subject matter that the Final Office Action asserts. Thus, Appellants respectfully submit that dependent claim 14 is allowable over Nassiff, Hon '801 and Hon '903, taken alone or in combination.

Dependent claim 15 provides the feature, to independent claim 7, wherein determining if the user's pronunciation deviated from existing pronunciations includes identifying, in the wave, the pronunciation of the corrected word. The Final Office Action asserted that Hon '903 teach this feature in column 7, lines 4-7. However, column 7, lines 4-7 of Hon '903 provides, "The training system would start out by prompting the speaker to pronounce various pre-selected words and then adapt the model accordingly." Appellants respectfully note that the cited portion of Hon '903 is discussing a training system, and thus, a user would not be correcting the words of dictated speech. Instead, the words are the known quantity to the training system, and the user's pronunciation of the known words is used to train the system. This is quite different than a method of learning with an automatic speech recognition system that includes detecting a change to dictated text. Accordingly, Appellants respectfully submit that not only does Hon '903 not provide the subject matter of dependent claim 15 that the Final Office Action alleges, but that one skilled in the art would not combine the training system of Hon '903 with the teachings of Nassiff and Hon '801. Further, Appellants respectfully submit that dependent claims 16-20 are allowable as well by virtue of their dependency, either directly or indirectly, from dependent claim 15.

3.3 Claims 21 and 22 are Allowable over Nassiff in view of view of Hon '801

Section Eight of the Final Office Action indicated that dependent claims 21 and 22 were rejected under 35 U.S.C. §103(a) based upon a Nassiff in view of Hon '801. On page 9 of the Final Office Action, the Examiner states that Hon '801 column 9, lines 36-40, column 7, lines 66-67, column 8, lines 1-3, and column 1, lines 33-36 and lines 54-56, disclose the claim 21 limitation of selectively learning from the nature of the correction includes adding at least one

word pair to the user's lexicon. Appellants respectfully submit that Hon ‘801 does not disclose the limitation at least because Hon ‘801 does not disclose a word pair or anything similar to a word pair.

Hon ‘801 column 9, lines 36-40 states:

> “If the unrecognized word is not in the lexicon 177, then the present invention stores the new word, along with predetermined attributes that the user provides, and assigns the word an initial unigram (step 181). The processing then returns to node A 141.”

Hon ‘801 column 7, lines 66-67 and column 8, lines 1-3 state:

> “If the unrecognized word is in the active lexicon of the program 75, then the language module adaptation 113 of the present invention is implemented. If the unrecognized word is not in the active lexicon, then the Add-to-Lexicon module 117 of the present invention adds the word to the lexicon.”

Hon ‘801 column 1, lines 33-36 states:

> “However, some of these errors stem from the fact that the spoken words are not in an active lexicon of the recognition program.”

Hon ‘801 column 1, lines 54-56 states:

> “Thus, there is a need for a method to reduce recognition error and rapidly adapt to unrecognized words in a speech recognition system.”

Appellants fail to see anything similar to a word pair in the above quoted passages or anywhere in Hon ‘801. On page 10 of the Final Office Action, the Examiner seems to indicate that the Hon ‘801 examples of “steep” and “step” are a word pair. The Hon ‘801 “steep” and “step” are not a word pair. They are an example of a misrecognized word and the correct word. Word pairs, among other things, help to prevent misrecognized pairs of words. An example of a word pair would be “too much” to prevent the word pair from being misrecognized as “two much.”

Since the cited references fail to disclose a word pair, Appellants fail to see how such references could disclose adding at least one word pair to the user’s lexicon. Appellants respectfully submit that claim 21 is allowable over Nassiff in view of Hon’ 801. Further,

Appellants respectfully submit that claims 22 and 23 are allowable as well by virtue of their dependency, either directly or indirectly from claim 21.

3.4 Claims 24 and 25 Are Allowable Over Nassiff in view of Gould

Section Twelve of the Final Office Action indicated that independent claim 24 and dependent claim 25 were rejected under 35 U.S.C. §103(a) as being unpatentable over Nassiff in view of Gould (EP 0773 532 A2). Section Twelve of the Final Office Action asserts that Nassiff provides the feature of claim 24 relative to inferring whether the change is a correction based at least partially upon the number of words changed. In this regard, the Final Office Action asserts,

> "Inferring whether the change is a correction (see col. 5, lines 60-61) based at least partially upon the number of words changed (e.g. It is obvious to the reference that the number of words are taken into consideration to find out which words were changed (see col. 5, lines 58-61, where replacement words and dictated words are one or more words. The deletion or typing over makes the inferring obvious in order to determine which words were edited or corrected.)"

However, this indicates that the Final Office Action has not given proper effect to the claim language. Specifically, the claim language does not recite determining which words were changed by considering the number of words. Instead, claim 24 recites inferring whether the change is a correction based at least partially upon the number of words changed. As set forth in Appellants' specification and as discussed in the Nassiff reference itself, it is important to understand whether replacement text represents correction of a misrecognition error rather than an edit. See Specification page 18, line 28 - page 19, line 2; and abstract of Nassiff. Independent claim 24 recites inferring the type of change based at least partially upon the number of words changed. Page 19 of Appellants' specification indicates that if the user changes a significant number of words in the dictated sentences, the user is probably editing based upon a change of mind. Accordingly, claim 24 is directed to determining the number of words changed, and using that information, at least partially, to infer whether the change is a correction as opposed to editing. Neither Nassiff nor Gould teach the utilization of such information for such an inference.

The Advisory Action mailed February 4, 2008 provides a response to the position. Specifically, the Examiner asserted that portions of Nassiff in column 5, lines 33-48 and column 5,

lines 50-61 show deletion and/or pasting being used as possible sources of user correction. The Examiner asserted,

> "If no overwriting occurs, then correction by the user is not determined to be made. In the Applicant's arguments, there is mention of determining an edit or a correction. However, the claim limitation in its current form does not mention anything about edits and further limitations from the specification are not read into the claims."

However, what dependent claim 24 does say is that the inference whether a change is a correction is based at least partially upon the number of words changed. This is a limitation in independent claim 24 that has been ignored by the Final Office Action. While Nassiff does discuss indicators that would inform the decision about the nature of the change, such indicator only include: whether the user has removed text immediately contiguous to the new word which has been inserted (column 5, lines 37-38); "if the backspace key or the delete key has been used to remove characters immediately contiguous to new text" (column 5, lines 41-43); and "if new text is inserted without overwriting dictated text" (column 5, lines 44-45). The other portion of Nassiff cited by the Advisory Action (column 5, lines 50-61) merely talks of the manner in which the replacement of a word may occur. Thus, while Nassiff does in fact provide indicators regarding whether a change is a correction, no indicators or inferences are based on the actual number of words changed. As set forth on page 19 of Appellants' specification, "If the user changes a significant number of words in the dictated sentences, the user is probably editing based upon a change of mind. Thus, a significant number of words being edited does not indicate a correction, but instead a change of mind. Using the number of words to infer whether a change is a correction is neither taught nor suggested in either the cited portions of Nassiff, nor the entire reference. Accordingly, Appellants continue to believe that independent claim 24 and dependent claim 25 are neither taught nor suggested by Nassiff and Gould, taken alone or in combination.

4. Conclusion: Claims 1, 3, 4, 6, 7 and 14-25 should be allowed.

In conclusion, Appellants respectfully submit that the rejection of claims 1, 3, 4, 6, 7 and 14-25 is improper, and that all claims 1, 3, 4, 6, 7 and 14-25 are in condition for

allowance. Accordingly, Appellants respectfully request that the Board reverse the rejection of claims 1, 3, 4, 6, 7 and 14-25 and find that such claims are allowable.

The Director is authorized to charge any fee deficiency required by this paper or credit any overpayment to Deposit Account No. 23-1123.

Respectfully submitted,

WESTMAN, CHAMPLIN & KELLY, P.A.

By:______________________________

Christopher R. Christenson, Reg. No. 42,413
Suite 1400, 900 Second Avenue South
Minneapolis, Minnesota 55402-3319
Phone:(612) 334-3222 Fax:(612) 334-3312

Appendix A: Claims On Appeal

Claims on appeal, as they currently stand:

1. (Previously Presented) A computer-implemented speech recognition system comprising:
    a microphone to receive user speech;
    a speech recognition engine coupled to the microphone, and being adapted to recognize the user speech and provide a textual output on a user interface;
    wherein the system is adapted to recognize a user changing the textual output and automatically, selectively adapt the speech recognition engine to learn from the change; and
    wherein the recognition engine is adapted to determine if a user's pronunciation caused an error, and selectively modify a probability associated with an existing pronunciation.

2. (Cancelled)

3. (Original) The system of claim 1, wherein the recognition engine includes a user lexicon, and wherein the user lexicon is updated if the correction is a word that is not in the user's lexicon.

4. (Previously Presented) The system of claim 1, wherein the recognition engine is adapted to selectively learn the user's pronunciation.

5. (Canceled)

6. (Previously Presented) The system of claim 1, wherein the recognition engine includes a user lexicon, and wherein the system is adapted to add at least one word pair to the user lexicon if the correction is not due to a new word, or a new pronunciation.

7. (Previously Presented) A method of learning with an automatic speech recognition system, the method comprising:

detecting a change to dictated text;

inferring whether the change is a correction, or editing;

wherein inferring whether the change is a correction, or editing includes comparing a speech recognition engine score of the dictated text and of the changed text;

if the change is inferred to be a correction, selectively learning from the nature of the correction without additional user interaction; and

wherein selectively learning from the nature of the correction includes determining if the corrected word exists in the user's lexicon, and if the corrected word does exist in the user lexicon, selectively learning the pronunciation.

8. (Canceled)

9. (Canceled)

10. (Canceled)

11. (Canceled)

12. (Canceled)

13. (Canceled)

14. (Previously Presented) The method of claim 7, wherein determining if the user's pronunciation deviated from existing pronunciations includes doing a forced alignment of a wave based on at least one context word if such word exists.

15. (Previously Presented) The method of claim 7, wherein determining if the user's pronunciation deviated from existing pronunciations includes identifying in the wave the pronunciation of the corrected word.

16. (Original) The method of claim 15, and further comprising building a lattice based upon possible pronunciations of the corrected word and the recognition result.

17. (Original) The method of claim 16, and further comprising generating a confidence score based at least in part upon the distance of the newly identified pronunciation with existing pronunciations.

18. (Original) The method of claim 16, and further comprising generating a confidence score based at least in part upon an Acoustic Model score of the newly identified pronunciation with existing pronunciations.

19. (Original) The method of claim 17, wherein selectively learning the pronunciation includes comparing the confidence score to a threshold.

20. (Original) The method of claim 19, wherein selectively learning the pronunciation further includes determining whether the new pronunciation has occurred a pre-selected number of times.

21. (Previously Presented) The method of claim 7, wherein selectively learning from the nature of the correction includes adding at least one word pair to the user's lexicon.

22. (Previously Presented) The method of claim 21, wherein the at least one word pair is added to the user's lexicon temporarily.

23. (Previously Presented) The method of claim 22, wherein the length of time the word pair is added to the user's lexicon is based at least partially upon the most recent time the word pair is observed and the relative frequency that the pair has been observed in the past.

24. (Previously Presented) A method of learning with an automatic speech recognition system, the method comprising:

detecting a change to dictated text;

inferring whether the change is a correction based at least partially upon the number of words changed; and

if the change is inferred to be a correction, selectively learning from the nature of the correction.

25. (Previously Presented) The method of claim 24, wherein if the change is inferred to be a correction, requesting a user confirmation.

Appendix B: Evidence

(None)

Appendix C: Related Proceedings

(None)